Project: Bank Marketing (Campaign)

Week 7 deliverables

Group name: Fleifel-solo

Name: Rania Tarek Fleifel

Email: raniatarekfleifel@gmail.com

Country: Egypt

College: Cairo university Faculty of engineering

Specialization: Data science

Internship Batch: LISUM13: 30

Submission date: 21st October 2022

Deliverable contents:

1) Problem description
2) Business understanding
3) Project lifecycle along with deadline
4) Data intake report
5) Git-hub Repo link

Data Glacier
Your Deep Learning Partner

Problem description

Provide ABC bank with a model that enables them to target costumers who're more probable to invest in their new term deposit product.

Business understanding

As ABC launches their new term deposit product, they need their outreach teams to effectively market the product to costumers whose interactions with the bank (loans, responsiveness to offers, etc) as well as their personal standing (job stability, marital status, age, etc) show high possibility of purchasing the deposit product. The need to focus on those costumers is -at the heart of it- the bank's strategy to effectively use the marketing team's resources to spread the deposit product among interested customers.

Deposit products promise costumers a high interest rate in return to locking an amount of their money for some time. Many factors decide whether a costumer invest in such product or not. The most important is his standing in life in general. For instance, customers who have savings beyond their day-to-day spending and have sitting-money would seemingly fit the profile of a perfect costumer. If a costumer's age is 60+, he's not a very good fit since he has limited resources and is retired, hence has no income except his pension which might not offer any excess to invest in this product. Customers who are used to taking loans could benefit as well if the interest rate from their savings covers their installments to the bank. Jobs play an important role as well in defining whether a costumer is a good fit. Doctors, engineers and similar prestigious jobs that are known to pay well are good candidates, as well as individuals with a long-standing job; 20+ work in a certain company shows stability.

Project lifecycle along with deadline

Week	Deadline	Milestone
07	19th October	Problem definition
08	26th October	Data inspection
09	2nd November	Data cleaning and transformation
10	9th November	Exploratory data analysis
11	16th November	EDA ppt (business pt-of-view)
12	23rd November	Model execution
13	30th November	Final delivery

Data Glacier
Your Deep Learning Partner

Data intake report

Name: Bank marketing dataset

Report date: 21st October 2022

Internship Batch: LISUM13: 30

Version: 1.0

Data intake by: Rania Tarek Fleifel

Data storage location: Bank marketing dataset (bank.zip)

Tabular data details:

1) bank.csv

Total number of observations	4521
Total number of files	1
Total number of features	17
Base format of the file	.csv
Size of the data	451 KB
Unique identifier feature	No unique identifier
Dupe validation	No duplicates

2) bank_full.csv

Total number of observations	45211
Total number of files	1
Total number of features	17
Base format of the file	.csv
Size of the data	4502 KB
Unique identifier feature	No unique identifier
Dupe validation	No duplicates


```
Information about bank.csv

Number of columns= 17

Columns names are ['age', 'job', 'marital', 'education', 'default', 'balance', 'housing', 'loan', 'contact', 'day',
'month', 'duration', 'campaign', 'pdays', 'previous', 'poutcome', 'y']

Number of entries= 4521.0

Size of file= 451 KB

Number of duplicated entries= 0

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4521 entries, 0 to 4520
Data columns (total 17 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   age        4521 non-null   int64
 1   job        4521 non-null   object
 2   marital    4521 non-null   object
 3   education  4521 non-null   object
 4   default    4521 non-null   object
 5   balance    4521 non-null   int64
 6   housing    4521 non-null   object
 7   loan       4521 non-null   object
 8   contact    4521 non-null   object
 9   day        4521 non-null   int64
 10  month      4521 non-null   object
 11  duration   4521 non-null   int64
 12  campaign   4521 non-null   int64
 13  pdays      4521 non-null   int64
 14  previous   4521 non-null   int64
 15  poutcome   4521 non-null   object
 16  y          4521 non-null   object
dtypes: int64(7), object(10)
```

```
Information about bank-full.csv

Number of columns= 17

Columns names are ['age', 'job', 'marital', 'education', 'default', 'balance', 'housing', 'loan', 'contact', 'day',
'month', 'duration', 'campaign', 'pdays', 'previous', 'poutcome', 'y']

Number of entries= 45211.0

Size of file= 451 KB

Number of duplicated entries= 0

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 45211 entries, 0 to 45210
Data columns (total 17 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   age        45211 non-null  int64
 1   job        45211 non-null  object
 2   marital    45211 non-null  object
 3   education  45211 non-null  object
 4   default    45211 non-null  object
 5   balance    45211 non-null  int64
 6   housing    45211 non-null  object
 7   loan       45211 non-null  object
 8   contact    45211 non-null  object
 9   day        45211 non-null  int64
 10  month      45211 non-null  object
 11  duration   45211 non-null  int64
 12  campaign   45211 non-null  int64
 13  pdays      45211 non-null  int64
 14  previous   45211 non-null  int64
 15  poutcome   45211 non-null  object
 16  y          45211 non-null  object
dtypes: int64(7), object(10)
```

Git-hub Repo link: Data science project-Bank Marketing Campaign